The Inter-Therapist Reliability of the Slump Test

This study examined the inter-therapist reliability of the slump test. Six pairs of physiotherapists tested a total of 93 patients currently receiving treatment for lumbar and/or lower limb symptoms. Each pair performed two slump tests on patients during a normal clinical visit. The slump test was positive if the patients' symptoms were reproduced, and subsequently decreased with cervical extension. A second definition of positive slump required decreased symptoms and increased knee extension with cervical extension. The results indicated that the slump test has high inter-therapist reliability which is consistent with reliability findings for related clinical tests of pain.

KATHLEEN PHILIP

Kathleen Philip, B.App.Sc(Phty), Grad. Dip. Manip. Ther., is a Staff Physiotherapist at the Austin Hospital, Heidelberg, Victoria.

PAUL LEW

Paul Lew, Dip. Physio., Grad. Dip. Adv. Manip. Ther., is a sessional lecturer in the Department of Physiotherapy, La Trobe University, and a private practitioner in Moonee Ponds, Victoria. He is currently undertaking studies for Master of Science (Anatomy) at the University of Melbourne.

THOMAS A. MATYAS

Thomas Matyas, B.A.(Hons), Ph.D., is a Senior Lecturer in the Department of Behavioural Health Sciences, Lincoln School of Health Sciences, La Trobe University, Carlton Campus, Victoria.

This paper is based on a study completed in partial fulfilment of the first author's Graduate Diploma in Manipulative Therapy

Since it was first described by Maitland (1978, 1979) the slump test has become widely advocated for assessment of patients with spinal and lower limb pain (Magarey 1984) particularly those with apparent hamstring injuries (Bourke *et al* 1986, Kornberg 1985, Kornberg and Lew 1987). Indeed Maitland (1985) states that the slump test is mandatory in the assessment of patients with lumbar symptoms. He also recommends that it be used to examine the cervical and thoracic regions if the patient has a history of trauma, if there is reason to believe that the canal structures are involved or it is necessary to prove their movement is normal, or if treatment of other faulty stuctures ceases to produce the predicted improvement. Selection of the region to be treated and the technique to be employed is also influenced by the slump test results. In addition, when the slump test is positive initially it may be used as a reassessment criterion to monitor the patient's progress during and between treatment sessions (Maitland 1985). Clearly the ability of the slump test to provide a reliable finding needs to be established, both on successive occasions and by different therapists.

However, despite the widespread clinical use of the slump test in Australia (Bourke *et al* 1986, Maitland 1985), little study has been done to determine its reliability. Maitland's initial study (1978) of 25 normal subjects was to establish the normal pain and range of movement response to the test. Maitland (1979, 1985), and Magarey (1983) described variations and clinical applications of the slump test but did not test the slump reliability. Lew (1979) used the slump test to implicate mobility of the canal structures as the limiting factor during straight leg raising (SLR). Kornberg (1985) used the slump test to implicate the lumbar spine as the source of hamstring pain in athletes. However these studies did not give further information about the reliability of the slump test.

Massey (1982) gave some indication of slump test reliability when she checked intra-examiner and inter-examiner error. Massey repeated the slump test on four subjects on three different occasions and found a maximum variation of 5° in 87.5 per cent of cases. This seems to indicate that the slump test is reliable in an intra-examiner test-retest situation. Inter-examiner error was tested by two therapists performing slump tests on nine different subjects. A variation of 4° in 94.5 per cent of tests was found suggesting that the slump test might also give high reliability in an inter-therapist situation. These results suggested that the slump test was reliable in reproducing ranges of movement. However in clinical practice the slump test is judged positive or negative according to the reproduction of the patients' symptoms and their subsequent behaviour with cervical extension. Massey does not indicate whether the slump tests were consistently positive or negative with regard to reproduction of the patients' symptoms, and her use of equipment to maintain position during the performance of the test is not repeated in the normal clinical situation. Further her reliability data is drawn from a very small sample of patients and therapists. While her results are encouraging, the reliability of the slump test in the clinical situation is still undetermined.

Method

Subjects

Twelve qualified physiotherapists, working in six pairs (A-F), agreed to participate in the study. The pairs comprised two pairs of physiotherapists with additional postgraduate training in manipulative therapy (manipulative therapists), one pair of physiotherapists concomitantly undertaking the manipulative therapy graduate diploma, one pair consisting of a manipulative therapist and a physiotherapist, and two pairs of physiotherapists.

Patients

The patient sample was drawn from six private physiotherapy practices. The sample consisted of 93 patients receiving treatment for low back pain with or without referred symptoms or lower limb pain. It was emphasized that subjects should not be selected for the trial on the basis of an expected positive response to the slump test. Subjects were excluded only if: pain was severe or irritable; the symptoms were thought to originate from an unstable discogenic disorder; or if neurological changes were present which were of recent onset or were progressing.

Procedure

Each patient was independently tested by a pair of therapists during a regular visit. The patients tested in this trial were being treated by one of the therapists in each pair, except Pair F whose patients were selected by a third therapist. Thus in Pairs A to E one therapist had greater knowledge of the patients' condition. Normal objective assessment of the patient was performed. A data sheet which was then completed by the treating therapist, prior to the performance of the slump tests, noted the age and sex of the subject, the area of symptoms (marked on a body chart), the duration of the current bout of symptoms and past history of symptoms. The first slump test was performed by one of the therapists, who recorded a positive or negative slump test result on a recording sheet. The patient was instructed to lie supine for 5 minutes. Then the second slump test was performed and recorded by the other therapist on a separate recording sheet. The second therapist had access to the patient's data sheet but not to the slump test recording sheet completed by the first therapist. One member of the therapist pair performed the first slump test in 50 per cent of the patients and the second therapist did so in the other 50 per cent. When one of the therapists was the patient's treating therapist the same protocol was followed. For pair F a third therapist completed the data sheet for these patients and did not reveal the contents to either slump test assessor. Thus neither therapist in this pair had an opportunity to form expectations about the outcome of the slump tests on the basis of other case data.

The slump test was performed in the following way (assuming the left leg was tested first):

(a) The patient was seated on the edge of the bench so that the posterior aspects of the thighs were supported by the bench and the popliteal creases were at the edge of the bench. The therapist ensured that the hips were in neutral with regard to medial rotation and adduction. The patient was instructed to sit erect and the therapist ensured that the sacrum was vertical. The patient placed his hands behind his back (Figure 1A). The therapist stood at the left side of the patient.

(b) The patient was instructed to let his back slump into full thoracic and lumbar flexion (Figure 1B). The therapist maintained the patient's chin in a nutral position preventing head and neck flexion. The patient was asked to report any pain, and if present whether it was similar to their symptoms.

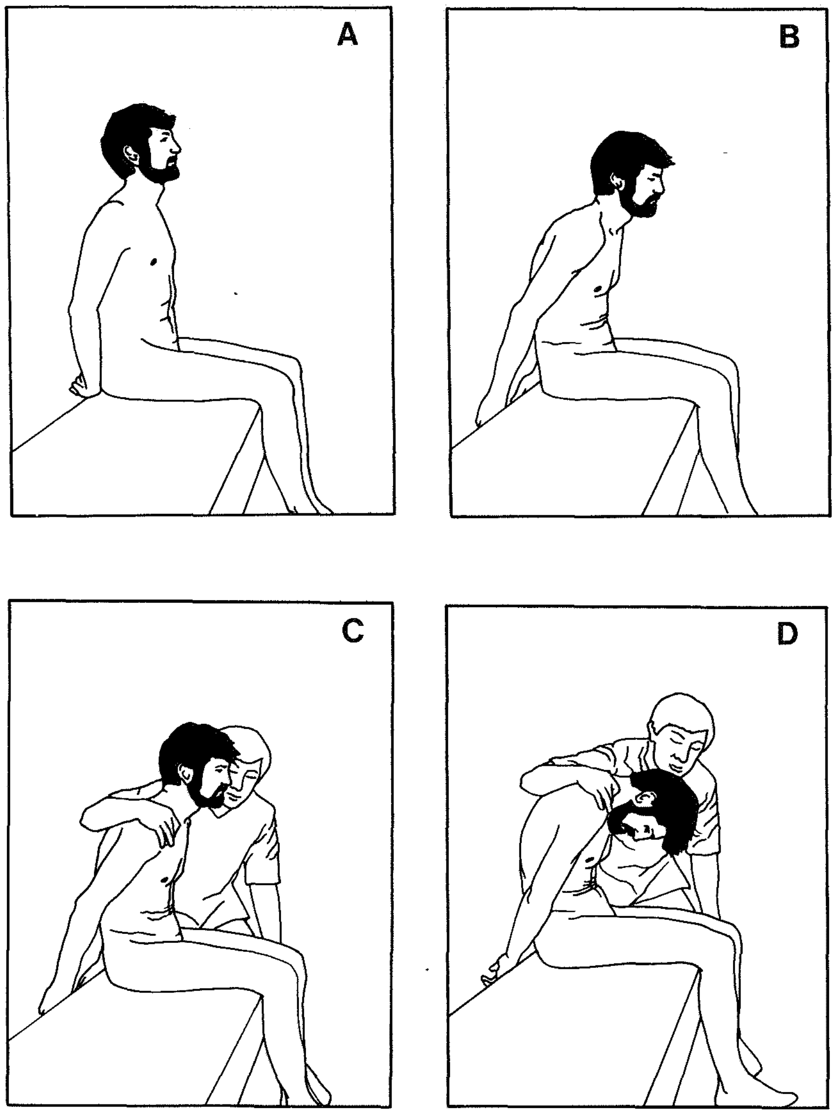
(c) The therapist's right arm and forearm applied overpressure to maintain the thoracic and lumbar spines in full flexion (Figure 1C). The patient was asked again to report any pain or discomfort and whether it was similar to their symptoms.

(d) Whilst the therapist maintained this overpressure the patient was instructed to fully flex his head and neck, taking chin to chest. The therapist applied overpressure to maintain this cervical position (Figure 1D). Again any reproduction of pain and its similarity to the patient's symptoms was noted.

(e) Using the left hand the therapist pulled the patient's left foot into maximum dorsiflexion.

(f) The patient was instructed to 'straighten your left knee as far as possible', the therapist emphasizing that the patient reach the point of maximum pain tolerance (P2) (Figure 1E). The patient was again asked if any pain was produced and if so whether it was similar to their symptoms. Steps (e) and (f) were repeated for the right leg and then for both legs together to ascertain which was the most provocative of their pain.

(g) Maintaining the most provocative position *eg* left knee extension,

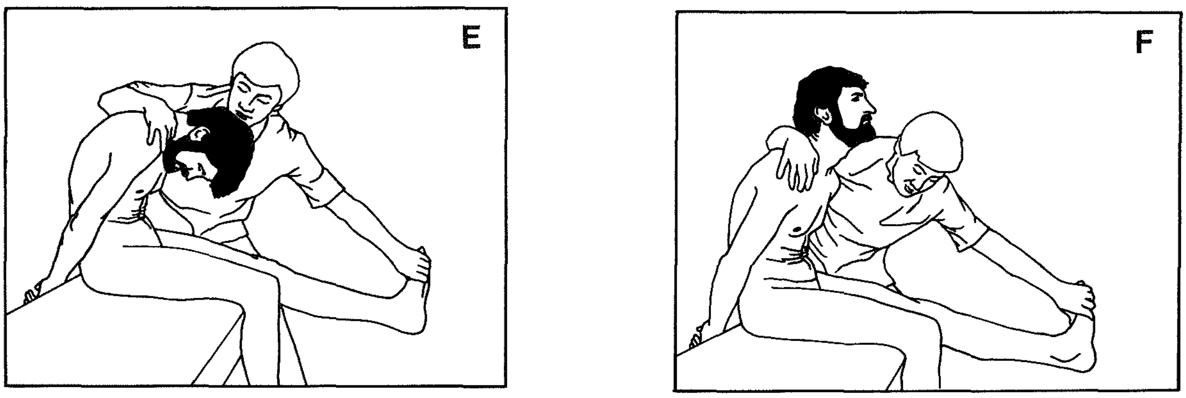


the therapist released the pressure to neck flexion taking care to maintain the same overpressure to the trunk flexion and knee extension and the patient was instructed to extend their neck (Figure 1F). The patient was asked 'What happens to your symptoms?' and whether he/she could extend the knee further.

(h) The results were recorded on a recording sheet. A positive slump test was recorded if part or all of the patient's symptoms were reproduced by the slump position and cervical extension caused the symptoms to decrease, with increased knee extension possible. A second type of positive slump test was recorded if the patient's symptoms were reproduced by the slump and cervical extension caused decreased symptoms but not increased range of knee extension. The second decision category, a negative slump test, was recorded if the slump did not reproduce the patient's symptoms, or if the patient's symptoms did not decrease with cervical extension.

Random checks on the adherence to the required procedure were conducted by the first author. These checks con-

Figure 1: *Sequence of subject postures in the Slump Test.*



sisted of visual appraisal of the performance of the slump test, the accuracy with which the data sheets were completed and the adherence to the test protocol.

Results

Results were analysed using Cohen's statistic kappa (Cohen 1960, Wallace and Elder 1980). Kappa (κ) expresses the proportion of obtained agreement relative to that expected to occur by chance (Matyas and Bach 1985). As the therapists' decisions were classified into two categories, either positive or negative, a high degree of random agreement was possible. Such random agreement might contribute to the apparent reliability of the test producing a false indication of intertherapist reliability. Kappa was deemed the most appropriate means of determining the observed reliability as it controls for the influence of such random agreement on the observed agreement scores.

Tables 1A and 1B present the raw obtained agreement rates for each pair (P_o), the agreement rate expected to occur by chance (P_e), calculated using the marginal distributions for each therapist and κ values for each pair of therapists. The tables also present P_o and P_e values averaged over all six pairs of therapists, and the overall κ based on these average P_o and P_e values. The main finding was a high level of agreement (mean κ= 0.89) between therapists when using the reproduction of symptoms and their subsequent decrease with cervical extension as the criterion of a positive slump test. Using the standard deviation of κ (Wallace and Elder 1980) we computed the 95% confidence interval for the mean value of κ, based on the full sample of 93 patients and 12 therapists. The confidence interval limits were from 0.81 to 0.97, supporting the conclusion of high reliability.

Statistically unsophisticated readers may wish to note that the kappa values for the particular therapist pairs presented in Table 1 should not be interpreted in the light of this confidence interval, which only applies to the overall sample of 93 patients. Confidence intervals for the particular therapist pairs would normally be wider given the larger standard errors inherent upon the smaller sample sizes which apply to particular therapist pairs. [$\sigma_\kappa = \sqrt{P_o(1-P_o)/N(1-P_e)^2}$ according to Wallace and Elder (1980)]. In addition, κ values of 1.0 present special mathematical difficulties because P_o values would be 1.0 for those cases (see above formula). Furthermore, the confidence interval estimates of κ values when true κ is very high should generally be regarded with caution because sampling distributions of statistics based on probabilities where the population value is extreme are asymetrical. The main point conveyed by the confidence in-

Table 1:
Frequencies of positive and negative diagnostic conjunctions and reliability statistics for two methods of clinical decision-making

1A: Results using symptom reproduction and subsequent decrease with cervical extension

Therapists		N	+/+	−/−	+/−	−/+	P_o	P_e	κ
Pair A	(PT/PT)	18	10	8	0	0	1.00	0.51	1.00
Pair B	(PT/PT)	14	6	6	1	1	0.86	0 50	0.72
Pair C	(MT/MT)	16	9	6	1	0	0.94	0.52	0.88
Pair D	(MT/MT)	10	4	6	0	0	1.00	0.52	1.00
Pair E	(MT/PT)	26	20	6	0	0	1.00	0.64	1 00
Pair F	(MTS/MTS)	9	6	2	0	1	0.89	0 54	0.76
Overall Values		93					0.94	0.53	0.89

1B: Results using symptom reproduction and decrease plus increased knee extension

Therapists		N	+/+	−/−	+/−	−/+	P_o	P_e	κ
Pair A	(PT/PT)	18	10	7	1	0	0.94	0.51	0 88
Pair B	(PT/PT)	14	6	6	1	1	0.86	0.50	0.72
Pair C	(MT/MT)	16	9	6	1	0	0.94	0.52	0.88
Pair D	(MT/MT)	10	4	6	0	0	1 00	0.52	1.00
Pair E	(MT/PT)	26	17	6	0	3	1.00	0.64	1.00
Pair F	(MTS/MTS)	9	6	2	0	1	0.89	0.59	0.73
Overall Values		93					0.92	0.54	0.83

Note. PT = Physiotherapist, MT = Qualified Manipulative Therapist, MTS = Student Manipulative therapist (postgraduate).

terval presented is to indicate that in general the inter-therapist reliability of the slump test was high.

A high degree of reliability (κ=0.83) was also obtained when a positive slump was defined as reproduction of symptoms, with cervical extension decreasing the symptoms and increasing the available range of knee extension. The slight apparent decrease in κ from 0.89 to 0.83 was not statistically significant. The 95% confidence interval for κ when a positive slump was defined as reproduction of symptoms with cervical extension decreasing the symptoms and increasing the available range of knee extension was 0.75 to 0.91.

Therapist variation in postgraduate qualifications did not appear to influence the results (Table 1A and B). When κ was calculated for the combined data of pairs comprising only manipulative therapists (pairs C and D) a value of 0.92 was obtained. The combined data of all remaining pairs yielded κ=0.87. The difference between these two values was not statistically significant.

Discussion

The results of this study indicated that the slump test has high reliability when the patients' symptoms are the criterion for positive or negative slump. This is in agreement with the previously demonstrated high reliability of SLR (Gajdosik and Lusin 1983, Hsieh *et al* 1983, MacFarlane 1981, Puentadura 1983) and forward flexion or fingertip-floor distance (Kwong 1981, Bruce 1981, Patterson 1982, Kippers and Parker 1987). In general, inter-pair variations in reliability were not large and the small sample size for specific pairs allows for substantial variation due to sampling error. Variations in therapist reliability according to qualifications were not consistent when pain change was the criterion. The observed disagreements seem unlikely to have been due to changes produced by the first slump test. Examination of the raw data revealed no observable trends of increased incidence of positive or negative results on the second slump test, which suggests that changes in the patients' conditions were not responsible for the disagreements.

Other possible reasons for discrepancies between therapists were differences in overpressure to full flexion, the subjective nature of P2 and perhaps individual tendencies towards detecting a positive or negative slump test. If one therapist produced consistently greater flexion movement with overpressure, and thus correspondingly greater tension, this therapist may have produced a positive slump, whereas a therapist who stretched the tissues to a lesser extent may have produced a negative slump. If this had occurred the results would have been expected to show a trend in the disagreements between therapists, one therapist showing consistently more positive slumps than the other. Although the small patient samples for each pair of therapists limit the conclusion, no such trends were evident. The number of positive and negative slumps seemed evenly distributed between therapists.

Individual tendencies among therapists towards calling a slump test positive or negative may influence the observed reliability. This should have produced a trend within disagreements if it was a significant factor, one therapist showing consistently more positive or negative slump tests than the other. This was not borne out by the results with the possible exception of Pair E.

As this study found high inter-therapist reliability for the slump test, it seems reasonable to suggest that intra-therapist test-retest reliability for this test will also prove to be high. From their review of studies assessing the reliability of pain and compliance Matyas and Bach (1985) found that intra-therapist reliability was consistently superior to inter-therapist reliability. This conclusion supports the use of the slump tst for clinical reassessment. However, this study did not attempt to measure the reliability of a scaled slump test, that is of a clinical decision based on changes in amount of pain produced, or changes in the degree of knee extension obtained. Our results indicate that slump is likely to be reliable for reassessment of change from positive to negative when that judgment is based upon the presence or absence of pain. However, the reliability of the slump test to detect other changes such as degree of change in pain or available range of movement is still an area of speculation. Massey (1982) found intra-therapist variations of 5° in 87.4 per cent of her small number of cases. Further study is necessary before the reliability of the slump test to detect therapeutic changes can be soundly supported.

The present study examined the slump test to P2, in concert with some other previous studies of the slump test. However examination to P2 has some potential difficulties, such as the stress on the patient and further irritation of the affected structures in patients with irritable conditions. In these cases investigations of the slump test to the onset of the reproduction of the symptoms (P1) is advisable. Assuming the same variables affecting P1 and P2 the present results suggest that a less stressful version of the slump test will also prove to be reliable, however this is yet to be shown.

Another interesting finding in this study was the large number of patients found to have positive slump tests (55 of 93 patients demonstrated positive slumps). Positive slumps were found in 12 of the 18 patients (66%) with lumbar pain only and in 37 (82%) of the patients with lumbar and lower limb symptoms. Six of the thirty patients (20%) presenting with only lower limb pain were found to have positive slumps. A positive slump test is thought to indicate a source of increased tension somewhere in the neuromeningeal tract (Maitland 1978). The anatomical extent of the neuromeningeal tract means that a large number of pathologies have the potential to increase tension in the tract. For example, the increased tension may arise from pathologies such as disc protrusion, which directly alter the state of the canal structures (Breig and Marions 1963). The tension may also increase because of nerve root adhesions (Fahrni 1966, Goddard and Reid 1965). Furthermore, increased tension may occur indirectly. Some possibilities include restricted intervertebral movement (Lew 1979, Massey 1982) or tight piriformis muscle (Cyriax 1982). Recently Klimiuk *et al* (1987) found that fibrinolytic defect occurred within two weeks of acute low back pain and that this fibrinolytic defect remain in the patients with persistent symptoms. They suggested that the failure to clear fibrin associated with the fibrinolytic defect can contribute towards the persistence of chronic inflammation and fibrosis. Chronic inflammation can hypersensitize the neuromeningeal structures to mechanical stress and fibrosis can reduce the mobility of the neuromeningeal structures thus predisposing them to adverse mechanical tension. Thus, the relatively large number of pathologies which could be responsible for increased neural tension may explain the high frequency of positive slump in this study. Some degree of bias in preselection of patients for the study cannot be completely dismissed as an explanation for the high incidence of positive findings. However, the high proportion of patients with lumbar and lower limb pain who demonstrated a positive slump test, and thus a presumed involvement of the canal structures, indicates the importance of these structures in lumbar pain conditions and the importance of this test in clinical assessment. Our findings are consistent with other recent studies (Bourke *et al* 1986, Kornberg 1985, Kornberg and Lew 1987). These have also obtained a surprising degree of positive slump tests in patients presenting with apparent hamstring injuries.

The high reliability reported in the present paper encourages further research on the slump test. The validity of the slump test as a test specific to the neuromeningeal tract needs confirmation. Such a finding taken together with the high reliability reported would permit the suggestion that the test should be employed as a screening device whenever it is reasonable to suspect the involvement of neural structures.

Acknowledgement

The authors wish to thank the sample of therapists in this study, who so generously donated their time and skills.

References

Bourke A, Alchin C, Little K and Sargood J (1986), Hamstring symptoms and lumbar spine relationship in sports people: a pilot study, *Proceedings of the Australian Physiotherapy Association National Conference, Hobart*, 309-321.

Breig A and Marions O (1963), Biomechanics of the lumbosacral nerve roots, *Acta Radiologica* (Stockholm), **1**, 1141-1160.

Bruce P (1981), The test-retest reliability of physiological movement as a method of assessing range of movement to the level of pain tolerance. Dissertation submitted in partial fulfillment for the Graduate Diploma in Manipulative Therapy, Lincoln Institute.

Cohen J (1960), A coefficient of agreement for nominal scales, *Educational and Psychological Measurement*, **20**, 37-46.

Cyriax J (1982), *Textbook of Orthopaedic Medicine, Vol. 1*, (8th ed.) Bailliere Tindall, London.

Fahrni W (1966), Observations on straight leg raising with special reference to nerve root adhesions, *Canadian Journal of Surgery*, **9** 44-48.

Gajdosik R and Lusin G (1983), Hamstring muscle tightness: Reliability or an active knee extension test, *Physical Therapy*, **63**, 1085-1088.

Goddard MD and Reid VD (1965), Movements induced by straight leg raising in the lumbosacral roots, nerves and plexus and in the intrapelvic section of the sciatic nerve, *Journal of Neurology, Neurosurgery and Psychiatry*, **28**, 12-18.

Hsieh C, Walker JM and Gillis K (1983), Straight-leg-raising test (Comparison of Three Instruments), *Physical Therapy*, **63**(9), 1429-1433.

Kippers V and Parker AW (1987), Toe-touch test, a measure of its validity, *Physical Therapy*, **67**(11), 1680-1684.

Klimiuk PS, Pountain GD and Jayson MIV (1987), Serial measurements of fibrinolytic activity in acute low back pain and sciatica, *Spine* **12**(9), 925-928.

Kornberg C (1985), Incidence of referred pain in Australian rules football players with a diagnosis of grade 1 hamstring strain. Dissertation submitted in partial fulfilment for Graduate Diploma in Manipulative Therapy, Lincoln Institute.

Kornberg C and Lew PC (1987), The effect of using slump as a stretching technique on grade one hamstring injuries, *Proceedings of the Fifth Biennial Conference of the Manipulative Therapists Association of Australia*, Melbourne, 183-191.

Kwong HF (1981), Test-retest reliability of pain onset assessed by active physiological movement. Dissertation submitted in partial fulfillment for Graduate Diploma in Manipulative Therapy, Lincoln Institute.

Lew PC (1979), The straight leg raise and lumbar stiffness. Dissertation submitted in partial fulfillment for the Graduation Diploma in Advance Manipulative Therapy, South Australian Institude of Technology.

MacFarlane A (1981), Test retest reliability of straight leg raise as determined by pain onset. Dissertation submitted in partial fulfillment for the Graduate Diploma in Manipulative Therapy. Lincoln institute.

Magarey M (1983), The use of slump in examination and treatment *Proceedings of the International Conference on Manipulative Therapy, Perth, MTAA 1983.*

Magarey M (1984), Canal signs: their significance in examination and treatment of the spine. Lecture presented at the School of Physiotherapy, Lincoln Institute.

Maitland GD (1978), Movement of pain sensitive structures in the vertebral canal in a group of physiotherapy students, *Proceedings of the Inaugural Congress of Manipulative Therapists Association of Australia, Sydney.*

Maitland GD (1979), Negative disc exploration: positive canal signs, *The Australian Journal of Physiotherapy*, **25**(3), 129-134.

Maitland GD (1985), The slump test: examination and treatment, *The Australian Journal of Physiotherapy*, **31**(6), 215-219.

Massey A (1982), The slump test: an investigation of the movement of pain sensitive structures in the vertebral canal in subjects with low back pain. Dissertation presented in partial fulfillment of the Graduate Diploma in Advanced Manipulative Therapy, South Australian Institute of Technology.

Matyas TA and Bach TM (1985), The reliability of selected techniques in clinical arthrometrics, *The Australian Journal of Physiotherapy*, **31**(5), 175-199.

Patterson S (1982), The test-retest reliability of lumbar flexion when limited by pain Dissertation submitted in partial fulfillment for Graduate Diploma in Manipulative Therapy, Lincoln Institute.

Puentadura L (1983), The effects of trunk position on straight leg raise in normal subjects. Dissertation submitted in partial fulfillment for Graduate Diploma in Manipulative Therapy, Lincoln Institute.

Wallace CJ and Elder JP (1980), Statistics to evaluate measurement accuracy and treatment effects in single-subject research designs, *Progress In Behavior Modification*, **10**, 39-79.